

DOCUMENT RESUME

ED 191 856

TM 800 390

AUTHOR Shermis, Mark D.; Albert, Susan L.
 TITLE CAPSAS: Computer Assisted Program for the Selection of Appropriate Statistics.
 PUB DATE Apr 80
 NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).
 EDES PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS Computer Oriented Programs: Decision Making: Educational Research: *Online Systems: *Statistical Analysis
 IDENTIFIERS *CAPSAS Computer Program

ABSTRACT

A computer-assisted program has been developed for the selection of statistics or statistical techniques by both students and researchers. Based on Andrews, Klem, Davidson, O'Malley and Rodgers "A Guide for Selecting Statistical Techniques for Analyzing Social Science Data," this FORTRAN-compiled interactive computer program was assembled to: (1) provide a "real time" decision model for selecting a particular statistic or statistical technique: and (2) serve as a prototype for "real time" decision models in research and evaluation designs. With CAPSAS, the user has interactive access to short, non-mathematical explanations of technical terms and brief descriptions of key words embedded in a question. CAPSAS can be integrated with other software computer simulation and analysis programs. (Author/CP)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED191856

CAPSAS: COMPUTER ASSISTED PROGRAM
FOR THE SELECTION OF APPROPRIATE STATISTICS

Mark D. Shermis and Susan L. Albert
The University of Michigan

ABSTRACT

A computer assisted program has been developed for the appropriate selection of statistics or statistical techniques. Based on Andrews, Klem, Davidson, O'Malley and Rodgers (1974), A Guide for Selecting Statistical Techniques for Analyzing Social Science Data, this FORTRAN-compiled interactive computer program was assembled to: 1) provide a "real time" decision model for selecting a particular statistic or statistical technique, and 2) serve as a prototype for "real time" decision models in research and evaluation designs. Through familiarization with the computer-based decision model, the user should acquire two benefits: 1) greater knowledge of statistical terms, techniques, and manipulations, and 2) greater likelihood of success in the selection of a statistic or statistical technique appropriate for analysis of the data.

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

M. Shermis

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

TM 800390

CAPSAS: COMPUTER ASSISTED PROGRAM FOR THE
SELECTION OF APPROPRIATE STATISTICS¹

Mark D. Shermis and Susan L. Albert
The University of Michigan

Decision-making trees or statistical guides are commonly used by social science researchers to select appropriate statistics for analyzing specified data sets (e.g., Andrews, Klem, Davidson, O'Malley and Rodgers, 1974; Harshberger, 1977). Once the statistic or technique has been chosen, the researcher may then perform the analysis by any one of the numerous statistical software packages currently available (e.g., SPSS, SAS, MIDAS, OSIRIS, etc.).²

This presentation serves as the introduction to the development of the Computer Assisted Program for the Selection of Appropriate Statistics (CAPSAS). Based on Andrews, et. al. (1974), A Guide for Selecting Statistical Techniques for Analyzing Social Science Data, this FORTRAN-compiled interactive computer program was assembled to: 1) provide a "real time," interactive decision model for selecting a particular statistic or statistical

¹Thanks to Joyce Carretta and Bob Dedrick for their programming help. We would also like to express our appreciation to Terry Davidson, L. S. Collet, Nancy Shiffler, Paul Pintrich, Paul Stemmer, and Jeanne Martin for comments on earlier versions of this draft.

²This presumes that the researcher knows how to use whatever system(s) are locally available, and can make intelligent choices among each system.

technique, and 2) serve as a prototype for "real time," interactive decision models in research and evaluation designs.

Printed guides currently in use utilize the notion of linear programming by asking "decisive questions" and eliciting user responses. One disadvantage of most guides is that the user must know how to properly interpret each question as he/she encounters it. In addition, the user must often be familiar with a wide variety of statistical terms. While most guides reference where supplemental information can be found, explanations of terms and questions are typically missing. This can be an impediment to a researcher who has forgotten what some of the "old jargon" means or a student who is currently trying to learn that "new jargon."

With the development of an interactive computer program to help select appropriate statistics, the user can ask a variety of different questions prior to making a decision. With CAPSAS, the user has interactive access to short, non-mathematical explanations of technical terms and brief descriptions of key words embedded in a question.³ Additionally knowledge learned with the interactive system can make room for alternative uses of instructional time. It can also provide for better utilization of existing computer capabilities. Through familiarization with an interactive computer-based decision model, the user should: 1) develop a greater knowledge of statistical terms, techniques and manipulations, and 2) increase the likelihood of success in the selection of a statistic or technique

³CAPSAS does provide references for those techniques, concepts, or terms which may require more in-depth explanation.

appropriate for analysis of a particular set of data.

CAPSAS facilitates two decision-making operations. Used as a heuristic aid, a teacher of research design or statistics courses could set up hypothetical problems and have students use CAPSAS to select appropriate statistics for subsequent data analysis. For example,

An English instructor wished to compare three methods of practicing spelling. The dependent variable of interest was the number of correct answers on a 20 word spelling test. In order to control for sex differences, the instructor randomly selected 3 groups of students, with equal numbers of boys and girls in each group. Because the instructor had the Iowa Test of Basic Skills scores test available for all the students, he desired to use vocabulary achievement as a mediating variable.⁴

Students faced with such a problem can consult with CAPSAS to discover optimal analytic strategies within the constraints of their situation. A session for the uninitiated user of CAPSAS trying to solve the English instructor's problem of what statistical technique to use for analyzing his data might look something like this:

```
*RUN CAPSAS 3=*SOURCE* 6=*SINK*
```

```
*EXECUTION BEGINS
```

```
CAPSAS:COMPUTER ASSISTED PROGRAM FOR THE  
SELECTION OF APPROPRIATE STATISTICS.
```

⁴Thanks to L. S. Collet for this example.

THIS PROGRAM IS BASED ON ANDREWS, F.M., KLEM, L.,
DAVIDSON, T.N., O'MALLEY, P.M., RODGERS, W.L.,
"A GUIDE FOR SELECTING STATISTICAL TECHNIQUES FOR
ANALYZING SOCIAL SCIENCE DATA". ANN ARBOR:ISR,
1974.

TYPE 1 FOR INSTRUCTIONS AND CAUTIONARY COMMENTS
OR HIT RETURN TO CONTINUE.

1

CAPSAS IS INTENDED TO HELP A DATA ANALYST
SELECT STATISTICS OR STATISTICAL TECHNIQUES WHICH WILL
BE APPROPRIATE FOR THE PURPOSES AND CONDITIONS OF A
PARTICULAR ANALYSIS.

TO USE CAPSAS, START WITH THE FIRST QUESTION AND
CHOOSE ONE OF THE ANSWERS PRESENTED THERE, OR ASK FOR
EXPLANATIONS BY SIMPLY TYPING THE NUMBER ASSOCIATED
WITH THE ANSWER. IF EXPLANATIONS ARE REQUESTED,
CAPSAS WILL PROVIDE AN EXPLANATION, AND THEN RESTATE
THE QUESTION. CAPSAS WILL PROCEED TO ASK QUESTIONS
UNTIL IT HAS ARRIVED AT A STATISTICAL TECHNIQUE AND/OR
A STATISTICAL MEASURE AND/OR STATISTICAL TEST APPRO-
PRIATE TO YOUR SITUATION-IF ONE WAS KNOWN TO THE
AUTHORS.

IN MANY ANALYSIS SITUATIONS IT IS POSSIBLE TO MAKE
ALTERNATIVE DECISIONS ABOUT THE NATURE OF ONES
VARIABLES, RELATIONSHIPS, AND/OR GOALS, AND THESE MAY
RESULT IN THE SELECTION OF AN APPROPRIATE STATISTIC,
MEASURE, AND/OR TEST. IT IS ALWAYS POSSIBLE TO USE
TECHNIQUES WHICH REQUIRE LESS STRINGENT ASSUMPTIONS
THAN THE ONES ORIGINALLY CONSIDERED. FOR EXAMPLE,
MEASURES OR TESTS MAY BE USED WHICH ARE APPROPRIATE FOR
A WEAKER LEVEL OF MEASUREMENT, OR TECHNIQUES APPRO-
PRIATE FOR NON-ADDITIVE SITUATIONS MAY BE USED EVEN THOUGH
THE VARIABLES ACTUALLY FORM AN ADDITIVE SYSTEM. RECALL
ALSO THAT TWO-POINT NOMINAL VARIABLES, AND RANKS,
MEET THE DEFINITION OF INTERVALLY SCALED VARIABLES.

The user
requests
information
on how to
use the
program

CAUTIONARY COMMENTS

1. WEIGHTED DATA, SMALL SAMPLE SIZES, COMPLEX SAMPLE DESIGNS, AND CAPITALIZATION ON CHANCE IN FITTING A STATISTICAL MODEL ARE SOURCES OF POTENTIAL PROBLEMS IN DATA ANALYSIS. CAPSAS DOES NOT DEAL WITH THESE COMPLICATIONS. IF ONE OF THESE SITUATIONS EXISTS, CAPSAS SHOULD BE USED WITH CAUTION.
2. THE STATISTICAL MEASURES PRESENTED ARE DESCRIPTIVE OF THE SAMPLE BEING EXAMINED. FOR SOME STATISTICAL MEASURES, THE VALUE OBTAINED WILL ALSO BE A GOOD ESTIMATE OF THE VALUE IN THE POPULATION AS A WHOLE, WHEREAS THE OTHER STATISTICS MAY UNDERESTIMATE (OR OVERESTIMATE) THE POPULATION VALUE. IN GENERAL, THE AMOUNT OF BIAS IS RELATIVELY SMALL AND SOMETIMES THERE ARE ADJUSTMENTS WHICH CAN BE MADE FOR IT. THESE ADJUSTMENTS ARE DISCUSSED IN GOOD STATISTICS TEXTS. IF A STATISTIC IS A BIASED ESTIMATOR, IT IS SO INDICATED IN CAPSAS.
3. IN PRINCIPLE, A CONFIDENCE INTERVAL MAY BE PLACED AROUND ANY STATISTIC. METHODS OF DOING THIS ARE NOT INDICATED IN CAPSAS. FORMULAS FOR COMPUTING CONFIDENCE INTERVALS FOR COMMONLY USED MEASURES ARE GIVEN IN STANDARD TEXTBOOKS.
4. CAPSAS DOES NOT EXPLICITLY CONSIDER POSSIBLE TRANSFORMATIONS OF THE DATA SUCH AS BRACKETING, USING LOGARITHMS, RANKING, ETC. TRANSFORMATIONS MAY BE USED TO SIMPLIFY ANALYSIS OR TO BRING DATA INTO LINE WITH ASSUMPTIONS. (FOR EXAMPLE, IT IS OFTEN POSSIBLE TO TRANSFORM SCORES SO THAT THE TRANSFORMED SCORES CORRESPOND TO A NORMAL DISTRIBUTION, CONSTITUTE AN INTERVAL SCALE, OR RELATE LINEARLY TO ANOTHER VARIABLE.) OCCASIONALLY, IT MAY BE WISE TO ELIMINATE CASES WITH EXTREME VALUES.
5. COMMON ASSUMPTIONS FOR INFERENCES BASED ON TECHNIQUES USING ONE OR MORE INTERVALLY SCALED VARIABLES (PARTICULARLY WHEN THE INTERVALLY SCALED VARIABLE IS A DEPENDENT VARIABLE) INCLUDE THE FOLLOWING: FIRST, THAT THE OBSERVATIONS ARE INDEPENDENT, I.E., THE SELECTION OF ONE CASE FOR INCLUSION IN THE SAMPLE DOES NOT AFFECT THE CHANCES OF ANY OTHER CASE BEING INCLUDED, AND THE VALUE OF A VARIABLE FOR ONE CASE IN NO WAY AFFECTS THE VALUE OF THE VARIABLE FOR ANY OTHER CASE; SECOND, THAT THE OBSERVATIONS ARE DRAWN FROM A POPULATION NORMALLY DISTRIBUTED ON THE INTERVALLY SCALED VARIABLE(S); AND THIRD, IF MORE THAN ONE VARIABLE IS INVOLVED, THAT THE INTERVALLY SCALED VARIABLE(S) HAVE EQUAL VARIANCES WITHIN CATEGORIES OF THE OTHER VARIABLE(S), I.E., THERE IS HOMOGENEITY OF VARIANCE. BIVARIATE OR MULTIVARIATE NORMALITY IS ALSO SOMETIMES ASSUMED.

The user is cautioned concerning statistical assumptions

HOW MANY VARIABLES DOES THE PROBLEM INVOLVE?

TYPE 1 FOR ONE VARIABLE

TYPE 2 FOR TWO VARIABLES

TYPE 3 FOR MORE THAN TWO VARIABLES

TYPE 4 FOR EXPLANATIONS

3

The program begins

IS A DISTINCTION MADE BETWEEN DEPENDENT AND INDEPENDENT VARIABLES?

1-NO

2-YES

3-EXPLAIN

2

IS THERE MORE THAN ONE DEPENDENT VARIABLE?

1-YES

2-NO

3-EXPLAIN

2

DO YOU WANT TO USE A COVARIATE, I.E., TO
STATISTICALLY REMOVE THE LINEAR EFFECTS OF ONE (OR MORE)
VARIABLES FROM THE DEPENDENT VARIABLE?

1-YES

2-NO

3-EXPLAIN

3

THE USE OF A COVARIATE GIVES ONE A PRECISE
METHOD FOR MAKING ALLOWANCE FOR AN UNCONTROLLED
VARIABLE AND TO SET FORTH THE SAMPLING ERROR ADJUSTMENT
WHICH IS NEEDED IN TESTING THE STATISTICAL SIGNIFICANCE
OF THE DIFFERENCE BETWEEN 'CORRECTED' MEANS. THE METHOD
(ANALYSIS OF COVARIANCE OR ANCOVA) IS APPLICABLE WHEN-
EVER IT SEEMS DESIRABLE TO CORRECT A DIFFERENCE ON A
DEPENDENT VARIABLE FOR A KNOWN DIFFERENCE ON ANOTHER
VARIABLE WHICH FOR SOME REASON COULD NOT BE CONTROLLED BY
MATCHING OR BY RANDOM SAMPLING PROCEDURES. (MCNEMAR, P. 414)

An explanation
is requested
for "covariate"

DO YOU WANT TO USE A COVARIATE, I.E., TO
STATISTICALLY REMOVE THE LINEAR EFFECTS OF ONE (OR MORE)
VARIABLES FROM THE DEPENDENT VARIABLE?

1-YES

2-NO

3-EXPLAIN

1

DO YOU WANT TO IGNORE POSSIBLE INTERACTIONS INVOLVING THE COVARIATE?

1-YES

2-NO

3-EXPLAIN

3

AN INTERACTION IS A MEASURE OF THE EXTENT TO
WHICH THE CRITERION MEAN FOR A TREATMENT COMBINATION
CANNOT BE PREDICTED FROM THE SUM OF THE CORRESPONDING
MAIN EFFECTS. THERE CURRENTLY IS NO STATISTICAL TECHNIQUE
WHICH CAN ANALYZE INTERACTIONS INVOLVING A COVARIATE
(WINER, P. 318).

An explan-
ation is
requested
for "in-
teraction"

DO YOU WANT TO IGNORE POSSIBLE INTERACTIONS INVOLVING THE COVARIATE?

1-YES

2-NO

3-EXPLAIN

1

DO YOU WANT TO TREAT THE DEPENDENT VARIABLE AND THE COVARIATES(S)
AS INTERVAL AND THE INDEPENDENT VARIABLE(S) AS NOMINAL?

1-YES

2-NO

3-EXPLAIN

3

OOPS--WE HAVE NOT WRITTEN AN EXPLANATION FOR
THIS WORD OR PHRASE YET, BUT HOPE TO HAVE ONE SOON

Well, we're
not perfect!

DO YOU WANT TO TREAT THE DEPENDENT VARIABLE AND THE COVARIATES(S)
AS INTERVAL AND THE INDEPENDENT VARIABLE(S) AS NOMINAL?

1-YES

2-NO

3-EXPLAIN

1

THE STATISTICAL MEASURE YOU HAVE CHOSEN IS

COVARIANCE ANALYSIS

THE STATISTICAL TEST FOR THIS MEASURE IS THE F TEST

TYPE 1 FOR EXPLANATION OR HIT RETURN TO CONTINUE

TYPE 1 TO CONTINUE PROGRAM OR 2 TO STOP

2

EXECUTION TERMINATED

The program comes
to an appropriate
technique

In the above mode of operation, CAPSAS serves as a device for problem solving by systematically investigating a series of possible questions. Advantages of operating in this mode include:

- 1) the student is able to ask for explanations of unfamiliar terms,
- 2) the student is able to ask for explanations of unfamiliar techniques, and
- 3) references to statistics or techniques are contained in one program.

In CAPSAS's operational mode, the educational or social science researcher can formulate an analysis strategy either based on some preordained design or with a knowledge of the limitations with which he/she is faced. For instance, the example provided above could be an instructional example of the teacher's or a "real world" problem. As in the instructional mode, the researcher obtains in a systematic way, one hypothetical solution -- if it is known. However,

as was articulated in the cautionary comments (p. 4), we advise experimentation with the different "branches" of the "tree," especially if less stringent requirements can be utilized for decision-making.

We think that this program is important not only because of its potential to systematically offer solutions, but also because it can be integrated with other software computer simulation and analysis programs. Simulation of research or evaluation designs such as FEHR Practicum (Collet, 1972) can already be performed on the computer (interactively, by the way) as can statistical analyses. In other words, the researcher can now choose appropriate statistics or techniques, execute a research design, and complete a statistical analysis -- all with the aid of a computer. This capability is invaluable in terms of time and monetary savings. The next step will be to create a program which will assist in the formulation of research or evaluation designs -- our next project.